AD_____

Award Number: W81XWH-08-1-0403

TITLE: Exploring the pathogenic and therapeutic implications of aberrant splicing in breast cancer

PRINCIPAL INVESTIGATOR: Dr. Sandeep Dave

CONTRACTING ORGANIZATION: Duke University
Durham, North Carolina 27705-5014

REPORT DATE: July 2011

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1JULY 2011	Final Scientific Report	1JULY 2008-30JUNE 2011

4. TITLE AND SUBTITLE

Exploring the Pathogenic and Therapeutic Implications of Aberrant Splicing in Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-08-1-0403

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Sandeep Dave and Jun Zhu

E-Mail: jun.zhu@duke.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Duke University

Durham North Carolina 27705-5014

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In this proposal, we aimed to a) systematically monitor splicing variant profiles in breast cancer susceptibility genes and b) explore the role of alternative splicing in breast chemotherapy using a global strategy. In doing so, we hoped to identify and validate candidate splicing variants involved in tumorigenesis using deep sequencing and functional assays.

15. SUBJECT TERMS
Deep sequencing, alternative splicing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	12	19b. TELEPHONE NUMBER *(include area code)*

Exploring the Pathogenic and Therapeutic Implications of Aberrant Splicing in Breast Cancer
W81XWH-08-1-0403

PRINCIPAL INVESTIGATORS: Sandeep Dave, MD, Jun Zhu PhD, William D Foulkes MB PhD

Table of Contents

Introduction

In this proposal, we aimed to a) systematically monitor splicing variant profiles in breast cancer susceptibility genes and b) explore the role of alternative splicing in breast chemotherapy using a global strategy. In doing so, we hoped to identify and validate candidate splicing variants involved in tumorigenesis using deep sequencing and functional assays. This annual report summarizes our progress in the final year of this award (we were granted a no-cost extension).

Body:

The aims of our original proposal were as follows:

A. Systematically monitor splicing variant profiles in breast cancer susceptibility genes.
B. Explore the role of alternative splicing in breast chemotherapy using a global srategy.

This work was done collaboratively with Drs. William Foulkes and Jun Zhu.

Although the originally proposed bar-code and chemosensitivity approach was not feasible after the departure of the original PI, Dr. Jun Zhu, we were able to develop a complementary and powerful approach using allele-specific RNA-sequencing to pursue the same goals.

Broadly, we pursued two complementary approaches to identify a role for alternative splicing in the disease. First, we have successfully developed allele-specific RNA sequencing as a viable method for the identification of global profiles of alternative splicing in breast cancers and, potentially, other malignancies. Second, we have further developed the experimental methods needed to functionally validate the effects of observed splicing variants.

These approaches have generated a wealth of data that we describe below.

Sample selection:

Parallel to the efforts of our co-PI, we chose the following cases of breast cancer cell lines: HCC1937, SUM149PT, HCC3153, SUM1315, MBC647, MPC7105, MPC298, MPC600, MPC960, LY3, LY10, TMD8, BJAB and LY19 for our sequencing work. These were compared to 14 sequenced lymphoid cells as controls.

Identifying alternative splicing events through allele-specific RNA sequencing

RNA-sequencing has emerged as a powerful method for identifying gene expression and alternative splicing events[2]. We have designed a workflow to identify the alternative splicing events in breast cancers. This method has the additional advantage of being able to identify genetic variants and gene rearrangements.

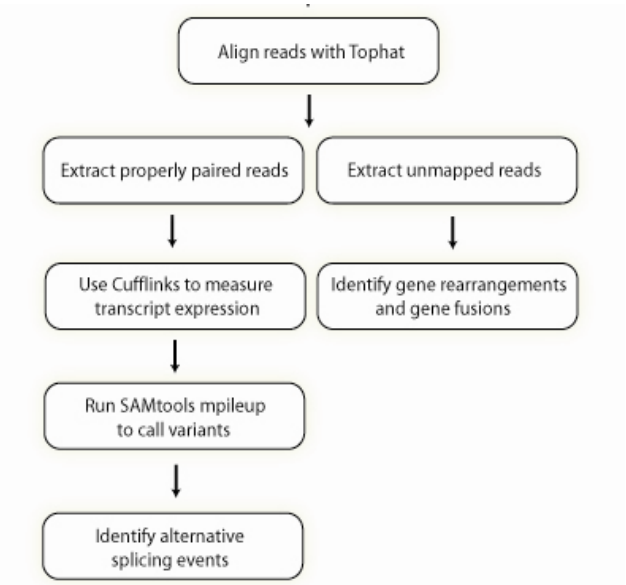

Figure 1: Workflow for RNA sequencing analysis

We have aligned RNA sequencing data from an these cases to the genome using Tophat[3]. After aligning the reads to the genome using Tophat, we run SAMtools mpileup[4] to call variants on all aligned reads. These variants will include those involved in alternative splicing events. We use the program deFuse[5] to search through all unaligned reads and look for those which span a gene fusion or gene rearrangement. Integrative Genomics Viewer was used to visualize the reads aligned to the genome[6].

On average, we found that 70% of the paired-end reads properly pair to the genome, which is consistent with recent transcriptome sequencing studies[7,8,9]. These sequencing reads map largely to annotated exonic regions and show evidence of alternative splicing (**Figure 2**). In all, we identified 248 such events that affected subgroups within our cases.

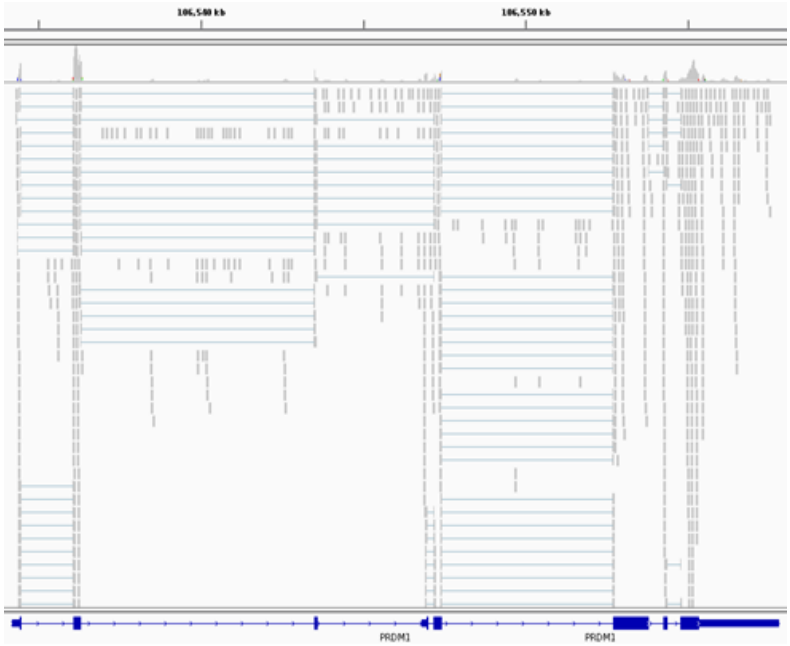

Figure 2: Reads aligned to PRDM1 using Tophat and Bowtie lie primarily in exons. Reads that span two exons are connected by a line across the intron that indicates splicing.

We also examined the strand-specificity of the sequencing reads and found strong evidence of strand-specificity. An example is shown in **Figure 3** on the genomic region that encodes the CD97 and DDX39 genes on overlapping loci on opposite DNA strands.

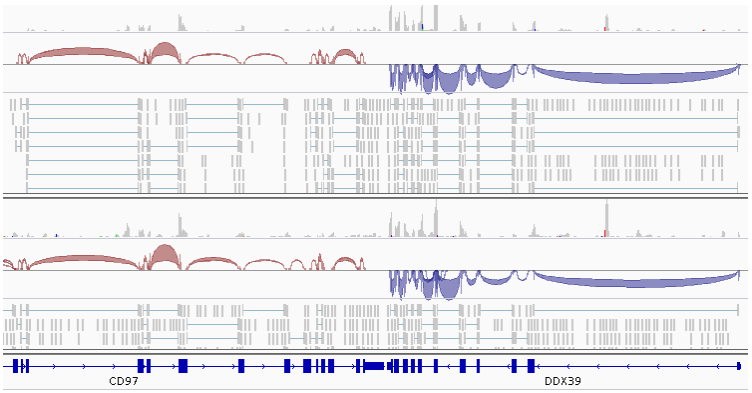

Figure 3: RNA-sequencing is allele (strand)-specific. CD97 and DDX39 are transcribed on opposite strands in close proximity to one another. Splicing is indicated in red on the positive strand and blue on the negative strand.

Identifying genetic variants in RNA sequencing data

Sequencing reads were processed according to the workflow described above. All variants that were called by SAMtools mpileup were filtered so that only those with a quality score of 30 or better (i.e. error of less than 1/1000) and a number of supporting reads (greater than 5) remained.

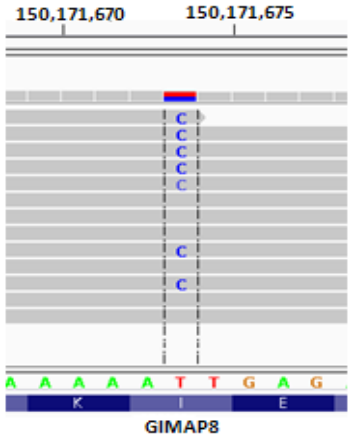

Figure 4: Variant in the gene GIMAP8 that is not present in the reference genome.

The genetic variants were annotated by genomic location using publicly available software SVA[10]. Variants occurring in known gene-coding regions will be further classified as synonymous, missense, nonsense, frameshift and splice side mutations. Separately, the total number of transitions and transversions were calculated for each sample. The deFuse[5] algorithm is being applied to discover all gene fusions and gene rearrangements in the transcriptome.

We estimate the rate of somatic mutations at ~1-2 x 10^{-6} and ~94% of these SNPs are already present in the dbSNP database[7]. An example of variant analysis is shown in figure 4, where we depict a missense variant in the GIMAP8 gene, which has been shown to be dysregulated in lung cancer. Its role in breast cancer remains to be defined.

Real-time polymerase chain reaction (RT-PCR) to measure validate deep sequencing data

In order to verify that our methods applying high throughput sequencing generated valid results, we applied real-time polymerase chain reaction (RT-PCR) in the same cases to

assess exon-level expression.1µg of RNA was reverse-transcribed with the ABI (Carlsbad, CA) High Capacity cDNA Reverse Transcription kit. Gene expression was measured with exon-spanning Taqman probes, and normalized to beta-2 microglobulin expression.

The primers used are shown in the table below.

Primer ID	Exon Number	Total Number of Exons	Ensembl Transcript ID	Sequence
LMO2_5b_F	5	6	ENST00000257818	GGAACCAGTGGATGAGGTGCTGCA
LMO2_5b_R	5	6	ENST00000257818	TCAGGCAGTCCTCGTGCCAGTACTG
LMO2_6b_F	6	6	ENST00000257818	CCCCCTTCCCAAGGCCTTAAGCTTTG
LMO2_6b_R	6	6	ENST00000257818	ACTCCTCCCCTCAAAATGAAGGTGTCT
STFB_4b_F	4	11	ENST00000519937	AAGTTCCTGGAGCAGGAGTGCAA
STFB_4b_R	4	11	ENST00000519937	AGTAGTCGATGACCAGGGGGAAGTAGT
STFB_8b_F	8	11	ENST00000519937	TGCCACCTCTGCATGTCCGTGA
STFB_8b_R	8	11	ENST00000519937	TTTCCCTGTCCAGCCAGGAGCCAA
LYPD6B_4b_F	4	7	ENST00000409642	CGCCCAGCACACAAGGTCAGCAT
LYPD6B_4b_R	4	7	ENST00000409642	GGTCGAGAGGAGGCCTCACATTATAGA
LYPD6B_6b_F	6	7	ENST00000409643	ACTTCACCAGCCACGGAAGAAGCA
LYPD6B_6b_R	6	7	ENST00000409644	TGTTCAGAATCTCGGCTGTGGTGGC

We further validated the accuracy of exon array measurement by performing quantitative RT-PCR to measure the exon-level expression of selected spliced genes (LMO2, SFTPB, LYPD6B) in the same samples (Figure 5).

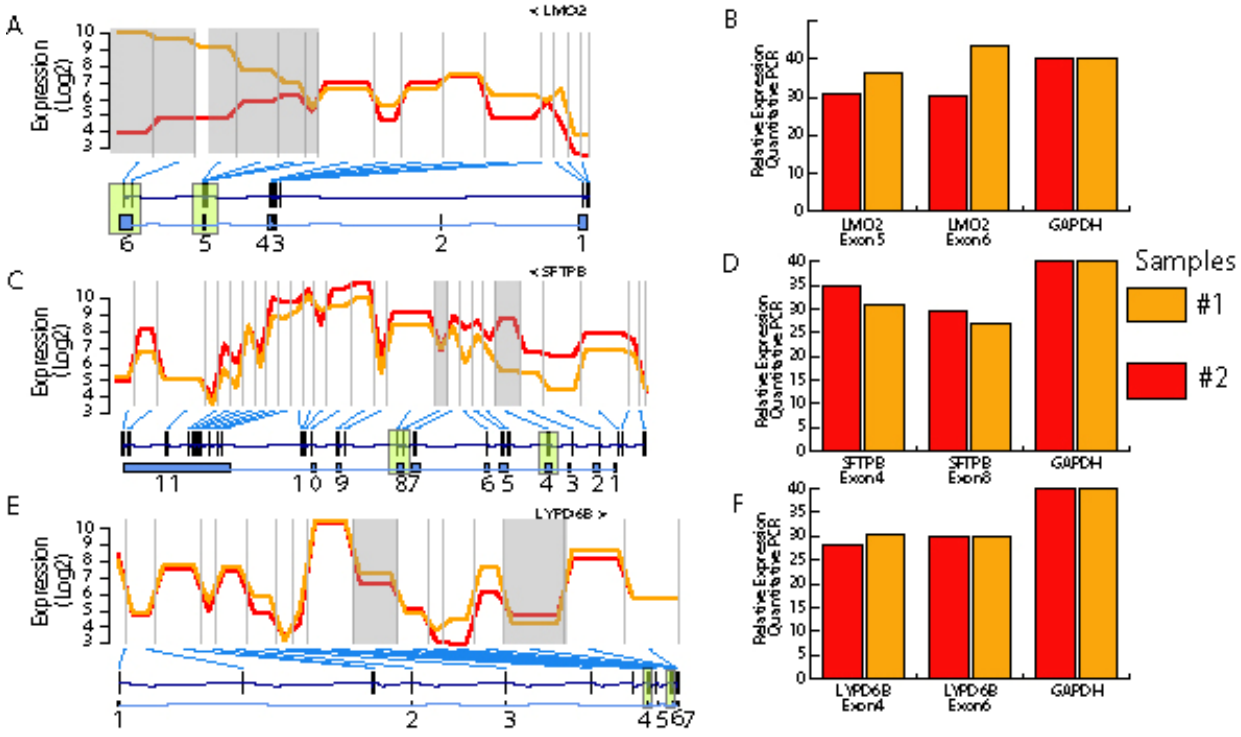

Figure 5: Validation of selected splicing events.
A The average expression levels of each exon in the LMO2 gene are depicted for different cases. Vertical gray bars indicate probesets mapped to the exon that were selected for validation by quantitative PCR. The blue connecting lines map the exons to Ensembl genes and the longest Ensembl transcript is shown to elucidate exon structure on the gene. The graph was generated using GenomeGraph[1].
B Quantitative real-time PCR to measure the expression of exons 5 and 6 of the LMO2 gene.
C The average expression levels of each exon in the SFTPB gene.
D Quantitative real-time PCR to measure the expression of exons 4 and 8 of the SFTPB gene.
E The average expression levels of each exon in the LYPD6B gene.
F Quantitative real-time PCR to measure the expression of exons 4 and 6 of the LYPD6B gene.

The identification of alternative splicing in LMO2 allows the identification of its functional role in cancer

In order to investigate the role of alternative splicing in the biology of oncogenes, we surveyed known oncogenes genes that showed highest levels of differential splicing. The LMO2 gene was selected as a suitable candidate because it was among the most highly differentially spliced genes. LMO2 is a known oncogene in the T cell leukemias[11-13] and is expressed highly in rapidly proliferating germinal center B cells[14].

The LMO2 gene comprises 6 exons. Our data indicate that only exons 5 and 6 of this gene are expressed at significantly higher levels in different cases (Figure 6A). Exons 5 and 6 encode the LMO2 LIM-binding domains that mediate protein-protein interaction. The remaining exons (exons 1-4) are non-coding, suggesting that alternative splicing of exons 5 and 6 represents the primary mechanism of LMO2 regulation in cancers.

We reasoned that the knowledge of the particular isoforms expressed in cancers provided an opportunity to discover the potential downstream targets of LMO2 and its function. We applied RNA-interference directed specifically at exon 6 of the LMO2 gene, which is alternatively spliced. Using lentiviral vectors to deliver two separate shRNA constructs to knock-down LMO2, we generated cell lines that stably expressed the shRNAs and had consequently lower levels of LMO2 mRNA (Figure 6B) and protein expression (Figure 6C).

We then performed gene expression profiling on the control cells and cells with LMO2-knockdown. We identified 338 genes that were down-regulated at least 1.5 fold in both RNA-interference experiments (Figure 6D). These genes are listed in Supplement Table 4. We performed gene set enrichment analysis[15] and found that genes related to proliferation[16] were highly associated with LMO2 knockdown (Figure 4E, $P<0.01$, $FDR<0.1$).

We also examined the effect of LMO2-knockdown on cellular proliferation rates. We found that LMO2-silencing was significantly associated with decreased cellular proliferation in in vitro experiments (Figure 6F). LMO2 silencing did not alter cellular viability (not shown) but significantly decreased the proliferation rate ($P<0.01$). These data indicate a potential role for LMO2 in mediating responsiveness to chemotherapy in tumors expressing the splicing variants that include exons 5 and 6.

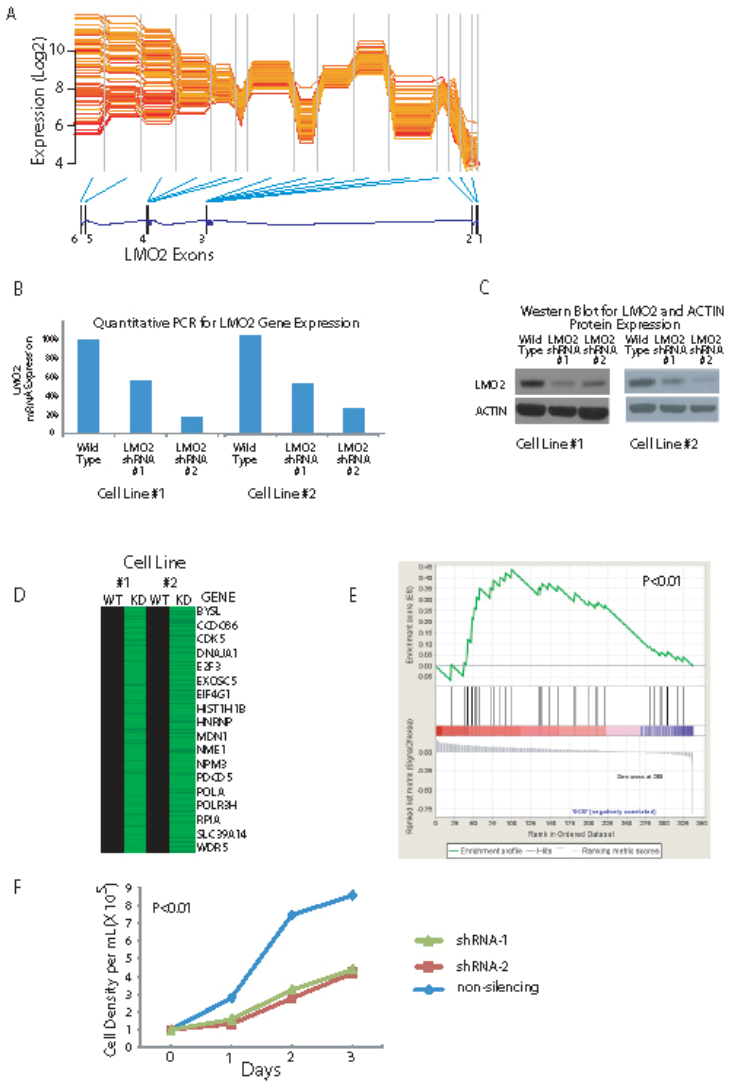

Figure 6: Splicing events in LMO2 result in altered phenotypes in cancer.
A Differential expression of exons 5 and 6 distinguishes different cancers. The remaining exons are expressed at roughly equivalent levels. Vertical gray bars indicate probesets mapped to the exon that were selected for validation by quantitative PCR. The blue connecting lines map the exons to Ensembl genes and the longest Ensembl transcript is shown to elucidate exon structure on the gene. The graph was generated using GenomeGraph[1].
B Lentiviral delivery of shRNAs targeting the 6th exon of LMO2 result in reduction of LMO2 mRNA levels compared to non-silencing controls. Two different shRNA constructs were used in two separate cell lines.
C Lentiviral delivery of shRNAs targeting the 6th exon of LMO2 result in reduction of LMO2 protein levels in the cell lines compared to ACTIN controls.
D Genes that were reduced at least 1.5-fold in response to stable knockdown of LMO2 gene in the cell lines are depicted. Down-regulated gene expression is shown in green over a 2-fold range. Selected genes are labeled in this heat-map.
E Gene set enrichment analysis demonstrates that proliferation is a key cellular process that is highly altered in response to LMO2-knockdown.
F LMO2-knockdown results in significant decreased cellular proliferation (P<0.01). Cellular viability was not significantly altered by LMO2-knockdown. Cell proliferation and viability of cells expressing the indicated shRNAs were determined by trypan blue exclusion.

These data point to a significant role for LMO2 in regulating cellular proliferation in cancer.

Summary of all work done during the granting period

Thus, although the originally proposed molecular barcode strategy could not be implemented after the departure of Dr. Jun Zhu from Duke, we have made significant findings in two major areas that are consistent with the original goals of the application:
1) We have developed allele-specific RNA-sequencing to identify alternative splicing events.
2) We have identified novel alternative splicing events in LMO2 and identified its downstream effects through RNA-interference experiments.

Key Research Accomplishments

• Development of allele-specific RNA sequencing as a novel approaches to analyze RNA Seq data
• Identification of novel alternative splicing events in breast cancers.
• Identification of alternative splicing as a novel mechanism for regulation of the oncogene LMO2.

Reportable Outcomes

Posters and Presentations

E. Lalonde Characterizating the BRCA1-deficient breast cancer transcriptomes by RNA-Seq [presentation]. Era of Hope Meeting, Orlando, Florida, August 4th 2011

Characterization of BRCA1-deficient breast cancer transcriptomes by RNA-Seq reveals novel transcript isoforms [poster]. American Society of Human Genetics Annual Meeting, Washington D.C., November 2-6 2010.

Published papers (work directly or indirectly supported by this award)

Lalonde E, Ha KCH, Wang Z, Bemmo A, Kleinman C, Kwan T, Pastinen T, and Majewski J (2011). RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. Genome Research, 21, 545-554.

Jima, D.D. et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. *Blood* 116, e118-27 (2010). (while this project is not solely related to breast cancer, the techniques were acquired during the 3 years of the DOD award)

Shuen AY, Foulkes WD. Inherited mutations in breast cancer genes--risk and response. J Mammary Gland Biol Neoplasia. 2011 Apr;16(1):3-15.

Martinez-Marignac VL, Rodrigue A, Davidson D, Couillard M, Al-Moustafa AE, Abramovitz M, Foulkes WD, Masson JY, Aloyz R. The effect of a DNA repair gene on cellular invasiveness: XRCC3 overexpression in breast cancer cells. PLoS One. 2011 Jan 24;6(1):e16394.

Conclusion

In this proposal, we have set out to evaluate the importance of splicing for breast cancer biology. Thus our work provides a starting point for the global identification of oncogenic events and the experimental methodologies for the functional characterization of such events. This work has produced a tremendous amount of data that will inform these questions for the next several years. We hope that our approach will lead to significant insights into the more general question of the importance of alternative splicing in breast cancer biology.

References

1. Durinck, S., Bullard, J., Spellman, P.T. & Dudoit, S. GenomeGraphs: integrated genomic data visualization with R. *BMC bioinformatics* **10**, 2 (2009).
2. Jima, D.D. et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. *Blood* **116**, e118-27 (2010).
3. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
4. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
5. McPherson, A. et al. deFuse: An Algorithm for Gene Fusion Discovery in Tumor RNA-Seq Data. *PLoS Comput Biol* **7**, e1001138 (2011).
6. Robinson, J.T. et al. Integrative genomics viewer. *Nat Biotech* **29**, 24-26 (2011).
7. Berger, M.F. et al. Integrative analysis of the melanoma transcriptome. *Genome Research* **20**, 413-427 (2010).
8. Maher, C.A. et al. Chimeric transcript discovery by paired-end transcriptome sequencing. *Proceedings of the National Academy of Sciences* **106**, 12353-12358 (2009).
9. Li, M. et al. Widespread RNA and DNA Sequence Differences in the Human Transcriptome. *Science* **333**, 53-58 (2011).
10. Ge, D. et al. SVA: software for annotating and visualizing sequenced human genomes. *Bioinformatics* **27**, 1998-2000 (2011).
11. Boehm, T., Foroni, L., Kaneko, Y., Perutz, M.F. & Rabbitts, T.H. The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. *Proc Natl Acad Sci U S A* **88**, 4367-71 (1991).
12. McCormack, M.P. & Rabbitts, T.H. Activation of the T-cell oncogene LMO2 after gene therapy for X-linked severe combined immunodeficiency. *The New England journal of medicine* **350**, 913-22 (2004).
13. McCormack, M.P. et al. The Lmo2 oncogene initiates leukemia in mice by inducing thymocyte self-renewal. *Science* **327**, 879-83 (2010).
14. Natkunam, Y. et al. The oncoprotein LMO2 is expressed in normal germinal-center B cells and in human B-cell lymphomas. *Blood* **109**, 1636-1642 (2007).
15. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
16. Shaffer, A.L. et al. A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunol Rev* **210**, 67-85 (2006).